DIE DIGITALE VERANTWORTUNGSLÜCKE: VORSCHLÄGE ZUR HAFTUNG FÜR ALGORITHMISCHES FEHLVERHALTEN

*Anna Beckers & Gunther Teubner**

1. DIE DIGITALE VERANTWORTUNGSLÜCKE: DREI FÄLLE 2
1.1. ROBO-ADVICE 2
1.2. PANAMA-PAPERS 2
1.3. FLASH CRASH 3
2. MASCHINENVERHALTEN, SOZIO-DIGITALE INSTITUTIONEN, HAFTUNGSRECHT 3
2.1. MASCHINENVERHALTEN UND SOZIO-DIGITALE INSTITUTIONEN 3
2.1.1 Wider den technikdeterministischen Kurzschluss 3
2.1.2 Die konstitutive Rolle der Sozialwissenschaften 5
2.2 SOZIO-DIGITALE INSTITUTIONEN UND HAFTUNGSRECHT 6
2.2.1 Digitale Assistenz 7
2.2.2 Digitale Hybridität 9
2.2.3 Digitale Interkonnektivität 10
2.3 HAFTUNGSSUBJEKTE 10
2.3.1 Nutzer/Betreiber 11
2.3.2 Netzwerkmitglieder 11
2.3.3 Industriesektor 12
2.4 RECHTSSTATUS DER ALGORITHMEN 12
2.4.1. Teilrechtsfähigkeit 13
2.4.2. Mitgliedschaft im Hybrid 13
2.4.3. Teil des Risikopools 13
3. DREI HAFTUNGSREGIMES 13
3.1. SYNOPSIS 13
3.2. HAFTUNGSRECHTLICHE NORMEN 14
3.2.1. Prinzipal-Gehilfen-Haftung 14
3.2.2. Netzwerkhaftung 14
3.2.3. Haftung des Fonds 15
4. DISKUSSION DER AUSGANGSFÄLLE 15
4.1. ROBO-ADVICE: DIGITALE AKTANTEN 15
4.2. PANAMA-PAPERS: DIGITALE HYBRIDE 16
4.3. FLASH CRASH: DIGITALE SCHWÄRME 17
5. AUSBLICK 17
BIBLIOGRAPHIE 19

Wenn autonome Algorithmen, die innerhalb unterschiedlicher sozio-digitaler Institutionen agieren, Fehlentscheidungen treffen, welche Konsequenzen ergeben sich für die rechtliche Haftung? Zur Entwicklung adäquater Haftungskonzepte rekurriert der Beitrag auf soziologische und philosophische Theoriestücke (1) der Personifikation nicht-menschlicher Akteure, (2) der Mensch-Maschine-Assoziation als emergentem sozialen System mit Qualitäten eines Kollektivakteurs und (3) der distribuierten Kognition in der Interkonnektivität von Algorithmen.

1. Die digitale Verantwortungslücke: Drei Fälle

1.1. Robo-Advice

Samathur Li Ki-kan, ein Tycoon aus Hongkong, verklagt Raffaela Costa, eine Investment-Brokerin, wegen des Fehlverhaltens eines Robo Advice Computers auf Ersatz eines Schadens von 23 Mio. Dollar. Der Supercomputer mit dem Namen K1 sollte Online-Quellen durchkämmen, um die emotionale Stimmung (!) der Anleger einzuschätzen und auf dieser Basis Prognosen für US-Aktientermingeschäfte abzugeben. Obwohl die ersten Simulationen vielversprechend erschienen, gingen beim realen Einsatz im Aktienhandel regelmäßig große Geldsummen verloren. Am 14. Februar 2018 verlor Li aufgrund einer Stop-Loss-Order über 20 Millionen Dollar (zum Fall Beardsworth und Kumar 2019; Pratt 2018).

Dies ist der erste bekannte Fall, in dem Schadensersatzforderungen für die Entscheidungen von autonomen Algorithmen wegen der von ihnen verursachten Anlageverluste vor Gericht gestellt wurden. In diesem Fall geht es um ein Blackbox-Problem: Wenn Menschen die Entscheidungen des Algorithmus nicht nachvollziehen können, wer haftet, wenn etwas schiefgeht? Nach bisher geltendem Recht gibt es keinerlei Schadensersatz, wenn die beteiligten menschlichen Akteure ihre Verhaltenspflichten erfüllt haben. Für bloßes Maschinenversagen haftet niemand. Dies ist die erste große Verantwortungslücke.

1.2. Panama-Papers

Im Investigativen Journalismus ist es schon heute üblich, Algorithmen zu nutzen. Dies gilt für die Recherche, insbesondere die Analyse einer großen Zahl von Dokumenten, aber auch für das Erstellen von journalistischem Content. So auch in diesem (hypothetischen) Falle, in dem ein internationales Konsortium von Journalisten autonome Algorithmen in einer komplexen Untersuchung einsetzte, um illegale Steuerpraktiken in einer Steueroase aufzudecken.[1] Die Schwierigkeit der Ermittlungsarbeit bestand darin, eine enorm große Anzahl von überaus komplexen Dokumenten zu analysieren. Algorithmen übernahmen die Arbeit des Markierens, Kategorisierens und Auswählens der relevanten Texte. Die Menschen waren in enger Interaktion an ihrer Arbeit beteiligt (Diakopoulos 2019, S. 13 ff.). Auch die spätere Veröffentlichung der Arbeit und ihre Verbreitung in den Nachrichten wurde von Algorithmen in der Form von Priorisierungs- und Filterentscheidungen beeinflusst.

Der Fall zeigt, wie erfolgreich die Zusammenarbeit von Algorithmen und Journalisten sein kann, um verborgene Skandale aufzudecken. Ohne die Hilfe von Algorithmen wäre, allein aufgrund der Masse an Dokumenten, der Steuerskandal nie an die Öffentlichkeit gelangt. Doch eine solche Praxis birgt ein beträchtliches Schadenspotenzial. Wer haftet etwa, wenn bei einer solchen Untersuchung Personen

---

* Für kritisch-konstruktive Hinweise danken wir Michael Heinlein. Auch aus den Diskussionen der KiMeGe-Gruppe haben wir vielfältige Anregungen erhalten.

[1] Dies ist ein fiktiver Fall, der aber öffentlich zugängliche Informationen zur Recherche der Panama-Papiere so aufbereitet, um die emergenten Eigenschaften einer Mensch-Algorithmen-Assoziation und ihre Rechtsprobleme zu illustrieren, hierzu im Detail: Panama Papers, The Journalists behind the Leak, 25. April 2016, abrufbar unter https://www.sueddeutsche.de/politik/panama-papers-the-journalists-behind-the-leak-1.2966929 (zuletzt besucht am 14.07.2022).

oder Unternehmen beschuldigt werden, die aber in Wahrheit nicht beteiligt waren? Besonders schwierig ist dies dann, wenn nicht eindeutig zu bestimmen ist, ob dem Algorithmus der Fehler unterlaufen ist oder ob menschliches Versagen vorliegt oder aber ob gerade die Zusammenarbeit zwischen Mensch und Algorithmus die falsche Anschuldigung verursachte.? Für die Haftung einer Mensch-Maschinen-Assoziation kennt das geltende Recht keine Grundlage. Hier klafft die zweite bedeutende Verantwortungslücke.

1.3. Flash Crash

Der algorithmische Hochfrequenzhandel ist der bekannteste Fall für Schäden, die durch die Vernetzung von Algorithmen entstehen. Berühmt wurde ein Flash-Crash auf dem US-amerikanischen Kapitalmarkt im Jahr 2010 (United States Commodity Futures Trading Commission und U.S. Securities & Exchange Commission 2016). Das US-Justizministerium machte einen Händler, Navinder Saro, für den Crash verantwortlich. Saro wurde wegen „Spoofing“ verklagt: Sein Algorithmus soll falsche Aufträge deswegen auf dem Markt platziert haben, um andere Handelsalgorithmen zu veranlassen, seinem Beispiel zu folgen. Dies erlaubte ihm, seine eigene Strategie in die genau entgegengesetzte Richtung zu ändern und damit Gewinne zu erzielen. Die Ermittlungen ergaben, dass das von Menschen initiierte Spoofing in der Tat eine Ursache für das Verhalten der anderen Algorithmen war und letztlich den Crash auslöste. Es konnte jedoch nicht eindeutig bestimmt werden, an welcher Stelle innerhalb der algorithmischen Operationen die unrechtmäßige Handlung stattgefunden hatte, die zu den katastrophalen Folgen führte. Die Algorithmen hatten zwar den programmierten Anweisungen gemäß gehandelt und die Transaktionen entsprechend ausgeführt. Der immense Schaden schien jedoch vor allem dadurch begründet, dass bei diesem algorithmischen Hochfrequenzhandel Algorithmen eingesetzt werden, die sehr ähnlich programmiert sind und damit ihre Handlungen wechselseitig verstärken (Yadav 2016, S. 1080). Dies deutet auf ein Herdenverhalten hin, in dem sich individuell vorprogrammierte Entscheidungen in einem interdependenten Prozess akkumulieren und eben dadurch die katastrophalen Folgen auslösen. In solchen Fällen, in denen Algorithmen individuell korrekt handeln, aber aufgrund ihrer gegenseitigen Verstärkung Schäden verursachen, versagt das traditionelle Haftungsrecht völlig. Dies ist die dritte Verantwortungslücke.

2. Maschinenverhalten, sozio-digitale Institutionen, Haftungsrecht

2.1. Maschinenverhalten und sozio-digitale Institutionen

2.1.1 Wider den technikdeterministischen Kurzschluss

Um diese gravierenden Haftungslücken zu bekämpfen, wurden in der äußerst lebhaften Reformdiskussion bereits zahlreiche Vorschläge zur Verantwortung für Algorithmen und entsprechende Haftungsregelungen unterbreitet: Sie reichen von der Haftung der Betreiber für gefährliche Anlagen über eine digitale Assistentenhaftung bis zur Haftung der Algorithmen selbst. Viele Autoren machen bei ihren Vorschlägen allerdings einen interdisziplinären Kurzschluss. Viel zu direkt schließen sie von den technischen Eigenschaften von Computern auf deren Risiken und auf rechtliche Haftungsregeln. „Technologie determiniert rechtliche Haftung“ – mit einem solchen Argument bleiben sie zu einfachen Kausalmodellen und ebenso zu einfachen

normativen Schlussfolgerungen verhaftet (z.B. Konertz und Schönhof 2020). Wir schlagen demgegenüber ein komplexeres Modell vor, das die unterschiedlichen sozialen Kontexte des Computereinsatzes ins Zentrum haftungsrechtlicher Argumente rückt und die Rolle der Sozialwissenschaften zur Bestimmung eines adäquaten Haftungsregimes betont.

Unser Ausgangspunkt ist eine in der IT-Forschung entwickelte Typologie, die drei Arten des Maschinenverhaltens unterscheidet: individuell, kollektiv und hybrid (Rahwan et al. 2019) .

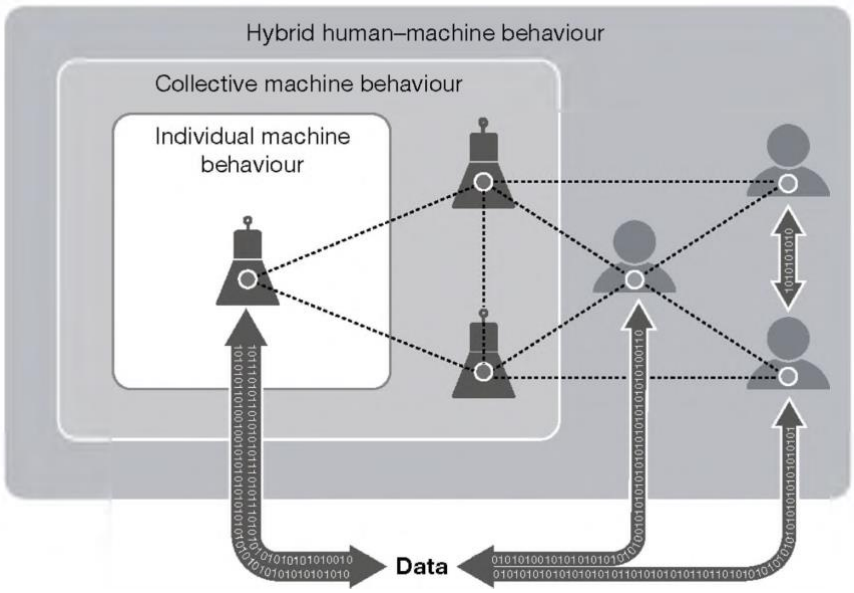

Abb. 1: Drei Arten des Maschinenverhaltens. Quelle: Rahwan et al. 2019, Abb. 1.

Diese Typologie dürfte für das Haftungsrecht von hoher Relevanz sein. Um aber den technikdeterministischen Kurzschluss zu vermeiden, ist es notwendig, „sozio-digitale Institutionen“ als intervenierende Variable zwischen Technik und Recht einzuführen. Unter sozio-digitalen Institutionen verstehen wir stabilisierte Komplexe sozialer Erwartungen, in unserem Falle Erwartungen in Bezug auf das Verhalten der Algorithmen in sozialen Kontexten. Solche Institutionen sind weder mit sozialen Systemen noch mit formalen Organisationen noch mit sozialen Beziehungen identisch. Vielmehr produzieren soziale Systeme, darunter auch formale Organisationen und natürlich auch zwischenmenschliche Beziehungen, mithilfe ihrer Kommunikationen Erwartungen, die sich, um eine klassische Formulierung zu bemühen, unter einer „idée directrice“ zu Institutionen verdichten [2]. Institutionalisiert sind solche Erwartungen, wenn Konsens zu ihrer Unterstützung unterstellt werden kann (Luhmann 1987 [1972], S. 64 ff.). Besonders wichtig ist in unserem Zusammenhang, dass Institutionen, weil sie aus Erwartungen bestehen, die Fähigkeit besitzen, Brücken zwischen unterschiedlichen Systemen und deren Erwartungen zu bauen (dazu genauer Teubner 2001, S. 360 ff.) Besonders auffällig ist ihr

[2] Während im „alten“ Institutionalismus (Santi Romano (1918), Carl Schmitt (1922), Hauriou (1933)) der Begriff der Institution zwischen System, Organisation und Normenensemble oszillierte, stimmen „neuer“ Institutionalismus und Systemtheorie trotz vieler Unterschiede in ihrem Verständnis von Institutionen als Erwartungskomplexe überein: North (1990), S. 3 f.; Luhmann (1965), S. 12 f.

Brückenschlag zum Recht, weil soziale Normen und Rechtsnormen die gleiche Wenn-dann-Struktur haben und sich nur in ihrer Ausrichtung auf unterschiedliche binäre Codes unterscheiden. In solchen Brückeninstitutionen treffen Erwartungen rechtlicher, wirtschaftlicher, politischer und technologischer Art aufeinander, und es ist häufig schwierig, zwischen den Erwartungen der beteiligten Systeme zu unterscheiden (Roberto Esposito 2022, S. 33 ff.)

Sozio-digitale Institutionen verdichten die noch unbestimmten „affordances" digitaler Technologien, also die kontingenten Nutzungschancen der Technologie, durch ihre gesellschaftliche Spezifikation zu konkreten Erwartungsbündeln (Graber 2021, 14). Sozio-digitale Institutionen integrieren diverse technische und soziale Erwartungen zu Chancen und Risiken des Einsatzes von Algorithmen in einem Prozess der Ko-Produktion.[3] Diese Institutionen dienen somit als wirksame strukturelle Kopplungen zwischen technischen und sozialen Systemen, darunter auch dem Rechtssystem. Gegenüber traditionellen sozialen Institutionen

> „zeichnen sich die instituierten und informellen Ordnungen der informationstechnologischen Kultur gerade dadurch aus, dass sie technisch werden. Die Codes und Programme, die technischen Standards und die mathematischen Algorithmen übernehmen jetzt einen Teil der Ordnungsleistung, den bislang ausschließlich symbolisch sinnhafte Ordnungen tragen mussten." (Vesting 2021, S. 220.)

Die Neuheit sozio-digitaler Institutionen gegenüber den bisher bekannten sozialen Institutionen besteht darin, dass sie auf der einen Seite die gewohnten an Bewusstsein gekoppelte Erwartungen bilden, auf der anderen Seite aber computergenerierte Erwartungen, die an rigide mathematische Kalkulationen gekoppelt sind. Diese asymmetrische „techno-digitale Normativität" schafft durchaus neue Gefährdungspotentiale (Karavas 2009, S. 478 ff.).

Erst die Einbettung in unterschiedliche sozio-digitale Institutionen erklärt die durchaus verwirrende Tatsache, dass Algorithmen manchmal als bloße Objekte oder Werkzeuge, manchmal als komplexe Personkonstrukte, manchmal als Mitglieder seltsamer Hybride, manchmal schließlich als gänzlich de-personalisierte Prozesse erscheinen. Zurecht macht Reyes 2021, 1483 ff. den jeweiligen sozio-technischen Kontext dafür verantwortlich, ob Algorithmen als „tools", als „conduits" oder als „autonomous entities" behandelt werden. Es gibt nicht die eine richtige Lösung für ihre Statuszuschreibung. Vielmehr hängen die unterschiedlichen Zuschreibungen von der Eigenrationalität und der Eigennormativität der jeweiligen sozio-digitalen Institutionen ab. Diese entscheiden darüber, wie unter den „affordances" der Technologie im jeweiligen sozialen Kontext gewählt wird und welcher Status dafür den Algorithmen zugewiesen wird:

2.1.2 Die konstitutive Rolle der Sozialwissenschaften

[3] Zur Ko-Produktion unterschiedlicher Sozialsysteme Feenberg (2017), 75; Jasanoff (2004); Winner (1980).

Damit plädieren wir für einen „institutional turn“ im Recht der Digitalität.[4] Erst im Kontext eigensinniger sozio-digitaler Institutionen wird vorstrukturiert, welche Formen sozialer Verantwortung und rechtlicher Haftung für das Handeln autonomer Algorithmen angemessen sind (Balkin 2015, S. 49; ähnlich Linardatos 2019.). Und genau für diese Zusammenhänge werden die Sozialwissenschaften als Vermittler zwischen IT-Wissenschaften und Rechtswissenschaft gebraucht. Der Einsatz sozialwissenschaftlicher Methoden „enables the identification of the common socio-legal needs that drive the personhood debates in the context of AI systems and demonstrates the importance of the socio-technical context for building appropriate legal rules related to AI“ (Reyes 2021, S. 1475). Diese Methoden sind in der Lage, die konkreten sozio-digitalen Institutionen und ihre Risiken mit ausreichender Tiefenschärfe zu analysieren und mit ausreichender Dichte zu interpretieren. In einer solchen „institutional analysis“ (Selznick 1969) ist die Rolle der Sozialwissenschaften keineswegs darauf beschränkt, wie es Juristen häufig missverstehen, nur deskriptiv-empirisch bestehende soziale Normen zu erforschen. Vielmehr besteht ihr zentraler Beitrag gerade darin, die verschiedenen Sinndimensionen sozio-digitaler Institutionen eigenständig zu interpretieren und sie mit „institutional imagination“ (Unger 1996) in interpretativen, funktionalen oder normativen Analysen weiterzuentwickeln.

Sozio-digitale Institutionen korrelieren in ihren Unterschieden mit den drei oben genannten Arten von Maschinenverhalten: Individuelles Maschinenverhalten wird in der Institution der digitalen Assistenz genutzt. Hybrides Maschinenverhalten lässt in der Interaktion zwischen Mensch und Maschine die Institution einer Mensch-Maschine-Assoziation entstehen. Kollektives Maschinenverhalten hingegen findet in der technischen Welt statt, die nur indirekt mit der gesellschaftlichen Sphäre verbunden ist. Hier ist die Gesellschaft den unsichtbaren Maschinen und deren vernetzten Operationen ausgesetzt. Jeder dieser Kontakte schafft also eine je eigene sozio-digitale Institution.

2.2 Sozio-digitale Institutionen und Haftungsrecht

Für jede sozio-digitale Institution schlagen wir ein spezifisches Haftungsregime vor. Damit tragen wir der Einbettung des Rechts in konkrete gesellschaftliche Institutionen Rechnung. So betont der politische Philosoph Roberto Esposito 2022, (S. 79 f.), dass

> „... das Recht nur innerhalb der konkreten Ordnung seinen Entfaltungsraum finden kann. Nicht auf einen abstrakten Normenkomplex reduzierbar, ist es materiell in dem sozialen Gefüge verwurzelt, aus welchem es hervorgeht. Vielmehr als eine Ordnung zu konstituieren, bringt das Recht, indem es menschliche Beziehungen regelt, ein Geflecht von Relationen zum Ausdruck, das im sozialen Körper bereits vorhanden ist.“

Freilich entstehen die sozio-digitalen Institutionen der Online-Welt nicht einfach ex nihilo. Sie alle haben ihren Ursprung in seit Langem bestehenden sozialen Institutionen der Offline-Welt. Gleichzeitig aber werden die etablierten Institutionen

[4] Dies folgt der Forderung nach einer institutionellen Wende in der Vertragsauslegung, die insbesondere für entstehende Institutionen in der digitalen Sphäre relevant wird, Wielsch (2018). Zum „institutional turn“ in den Sozialwissenschaften Roberto Esposito (2022), 33 ff.

durch die „Invasion“ der Algorithmen grundlegend verändert. Jede sozio-digitale Institution weist eigene neuartige Schädigungsrisiken auf: Das Risiko *digitaler Assistenz* besteht darin, dass Aufgaben statt an Menschen an autonome Algorithmen delegiert werden, deren Entscheidungen nicht mehr kontrollierbar sind. *Mensch-Maschine-Assoziationen* erzeugen das Risiko emergenter Kollektiventscheidungen, die nicht auf Individualentscheidungen der beteiligten Algorithmen oder Menschen zurückführbar sind. Das Risiko *digitaler Interkonnektivität* besteht darin, dass gesellschaftliche Kommunikation einer undurchschaubaren Vernetzung von interagierenden Algorithmen ausgesetzt ist (zu Einzelheiten Beckers und Teubner 2021, S. 14 ff., 45 ff., 90 ff., 111ff.).

Als vermittelnde Disziplin zwischen Informatik und Rechtswissenschaft dürften die Sozialwissenschaften über die analytischen Instrumente verfügen, die ko-evolutionäre Dynamik zwischen technologischen Innovationen, sozio-digitalen Institutionalisierungen und haftungsrechtlichen Entwicklungen zu analysieren (Nassehi 2019, S. 15 ff.). In der Wahl zwischen relevanten sozialwissenschaftlichen Theoremen kann sich das Haftungsrecht dabei nicht, wie es jedoch vielen Autoren vorschwebt, ausschließlich auf ökonomische Analysen stützen. Diese konzentrieren sich auf Anreize für Sorgfaltsstandards und Aktivitätsniveaus, sind aber gegenüber umfassenderen gesellschaftlichen Problemen, insbesondere den Interessen der Opfer an einer Entschädigung oder ökologischen Schädigungen, relativ gleichgültig. Gemäß dem in der Philosophie und Soziologie entwickelten Prinzip der „Transversalität“ (Welsch 1996; Wagner 1996) versuchen wir deshalb, relevante Erkenntnisse auch aus anderen Sozialwissenschaften zu gewinnen. Hierbei rekurrieren wir insbesondere auf Theorien zur gesellschaftlichen Personifizierung von Algorithmen, zu emergenten Eigenschaften von Mensch-Algorithmus-Assoziationen und zur verteilten Kognition in miteinander vernetzten Algorithmen. Letztlich orientieren sich unsere Vorschläge zur Risikohaftung an einer Kombination ökonomischer und soziologischer Theoreme.

Weil die involvierten sozio-digitalen Institutionen ganz unterschiedlich sind, wird man von vornherein auf abstrakt-generelle Haftungsregeln verzichten müssen (Krafft et al. 2022, unter 1; Lohsse et al. 2019, S. 21 ff.). Wir schlagen stattdessen konkretere unterschiedliche Haftungsregimes vor, die der unterschiedlichen institutionellen Einbettung des algorithmischen Verhaltens Rechnung tragen. Die Haftungsregimes reagieren auf die je verschiedenen Risiken der drei sozio-digitalen Institutionen: auf das bisher unbekannte Risiko einer Delegation von Aufgaben an autonome Algorithmen in digitaler Assistenz, auf das Risiko einer neuartigen Kollektivität von hybriden Mensch-Algorithmus-Assoziationen und auf das ebenso bisher unbekannte Risiko, dass die Gesellschaft der undurchsichtigen Vernetzung von Algorithmen in digitaler Interkonnektivität ausgesetzt sind.

2.2.1 Digitale Assistenz

Individuelles Maschinenverhalten bezieht sich auf intrinsische Eigenschaften eines einzelnen Algorithmus, dessen Dynamik durch Quellcode oder Design in seiner Interaktion mit der Umwelt bestimmt wird (Rahwan et al. 2019, S. 481). Diese technischen Eigenschaften allein können jedoch nicht den Ausschlag geben, ob Algorithmen als autonome Akteure qualifiziert werden können oder nicht. Stattdessen bestimmen sozio-digitale Institutionen, ob Algorithmen den sozialen Status von bloßen

Instrumenten einnehmen, ob sie Agenten in Prinzipal-Agenten-Beziehungen sind oder ob sie – als eine mögliche zukünftige Entwicklung – zu unabhängigen sozio-ökonomischen Akteuren werden, die ihre Eigeninteressen verfolgen (‚E-persons').

Für Prinzipal-Agent-Beziehungen können sozialwissenschaftliche Theorien klären, unter welchen Bedingungen die Institution der digitalen Assistenz entsteht. Wenn die Aufgabendelegation von einem menschlichen Akteur (oder einer Organisation) an einen Algorithmus zwei autonome, aber interdependente Handlungsketten erzeugt, dann entsteht zwischen diesen eine Prinzipal-Agent-Beziehung (z. B. Krafft, Zweig und König 2022, unter 1). Solche Beziehungen setzen notwendigerweise sowohl für den Prinzipal als auch für den Agenten soziale Handlungsfähigkeit voraus. Daher wird innerhalb der sozio-digitalen Institution eine selektive Zuschreibung von Personalität an bestimmte digitale Prozesse erforderlich.

Personifizierung von Algorithmen – für diesen komplexen sozialen Prozess liefern mehrere soziale Theorien die entsprechende Begründung. Die Wirtschaftswissenschaften tragen zu diesem Thema relativ wenig bei. Wenn sie den Einsatz von Algorithmen in Märkten beobachten, setzen sie implizit Algorithmen als rationale Akteure voraus. Im Gegensatz zu engen Rational-Choice-Annahmen begreift die soziologische Theorie die Personifizierung als einen performativen Akt, der die soziale Realität eines Akteurs überhaupt erst begründet, der aber nicht auf eine spezifische Rationalität festgelegt werden kann (Muhle 2018; Beckers und Teubner 2021). Komplementär dazu definiert die Akteur-Netzwerk-Theorie die interaktiven Eigenschaften, die einen Algorithmus zu einem „Aktanten" machen. Sie zeigt außerdem auf, in welchen Aspekten sich dieser von einem menschlichen Akteur unterscheidet (Latour 2004, S. 62ff.). Die Informationsphilosophie legt fest, unter welchen Bedingungen algorithmische Handlungen als autonom gelten können (Floridi und Sanders 2011, S. 192ff.). Die Systemtheorie beschreibt detailliert, wie in einer Situation doppelter Kontingenz die emergente Mensch-Maschine-Kommunikation die soziale Identität des Algorithmus und seine Handlungsfähigkeit überhaupt erst konstituiert (Elena Esposito 2017). Dabei schafft jeder soziale Kontext für Algorithmen seine eigenen Kriterien der Personalität, die Wirtschaft nicht anders als die Politik, die Wissenschaft, die Moral oder das Recht. Konkrete Operationen in den verschiedenen sozialen Systeme schreiben den Algorithmen als ihren ‚Personen' auf unterschiedliche Weise Handlungen, Rechte und Pflichten zu und statten sie mit spezifischen Ressourcen, Interessen, Absichten, Zielen oder Präferenzen aus. Für eine solche Zuschreibung von Personalität ist kein voller Konsens der beteiligten Akteure erforderlich. Denn wie schon gesagt, beruhen Institutionalisierungen auf Konsensunterstellungen (Luhmann 1987 [1972], S. 64 ff.), die selbst bei hoher Meinungspluralität wirksam sind. In der politischen Philosophie wird detailliert beschrieben, wie die konkrete Übertragung der „potestas vicaria" die Stellvertreter-Personalität von Algorithmen konstituiert, die zugleich „eindeutige Risiken und Gefahren für die Modernität impliziert" (Trüstedt 2020, S. 196 f.; Trüstedt 2022, V.1).

Ergebnis sozialer Personifizierung ist, dass nun technische Risiken in soziale Risiken transformiert werden. Die Risiken stammen nicht mehr einfach von fehlgeschlagenen Kausalverläufen, sondern von fehlerhaften Entscheidungen der Quasi-Akteure, die Erwartungen über die Handlungen digitaler Aktanten enttäuschen. In Situationen der digitalen Assistenz ist nicht mehr eine instrumentelle Subjekt-Objekt-Beziehung wirksam, sondern eine Subjekt-Subjekt-Beziehung, genauer eine Prinzipal-

Agent-Beziehung mit ihren typischen kommunikativen Risiken. Je mehr sich die Institution digitaler Assistenz in den Online-Transaktionen stabilisiert, desto stärker wird das Recht herausgefordert, seinerseits nach eigenen Kriterien zu entscheiden, welchen Grad an Rechtspersönlichkeit es den digitalen Akteuren zubilligt. Deswegen sind Haftungsregeln erforderlich, die auf Handlungsrisiken digitaler Akteure und nicht bloß auf die Kausalrisiken gefährlicher Objekte reagieren. Genau aus diesem Grund halten wir die rechtspolitischen Vorschläge, eine neue Gefährdungshaftung einzuführen oder die Regeln zur Produkthaftung zu modifizieren, welche die Algorithmen schlicht als Objekte, als gefährliche Anlagen oder als fehlerhafte Produkte behandeln, für nicht adäquat. Sie ignorieren gerade das Neuartige an Algorithmen – die in der digitalen Assistenz zugeschriebene autonome Entscheidungsfähigkeit. Auf (fehlerhafte) Entscheidungen von Algorithmen im Kontext digitaler Assistenz sind vielmehr die Regeln der Prinzipal-Gehilfen-Haftung anzuwenden. Der Prinzipal ist gebunden, wenn der Algorithmus als Stellvertreter Verträge abschließt, und er haftet, wenn der Algorithmus als Gehilfe eine rechtswidrige Entscheidung trifft und einen Schaden verursacht.

2.2.2 Digitale Hybridität

Im Vergleich zur digitalen Assistenz können die Sozialwissenschaften zum Verständnis der hybriden Mensch-Maschine-Assoziationen, d.h. der eng verflochtenen Interaktionen von Algorithmen und Menschen, auf andere Weise beitragen. Würde man hier mit dem individualistischen Ansatz der Prinzipal-Agenten-Beziehung versuchen, den Verantwortungsbeitrag einzelner menschlicher oder algorithmischer Handlungen zu identifizieren, so übersähe man, dass sich hier ein neues Kollektiv etabliert hat – die Mensch-Maschine-Assoziation. Sie entwickelt Kollektivrisiken, die sich qualitativ von den Individualrisiken der digitalen Assistenz unterscheiden. Während die digitale Assistenz mit den Risiken einer neuartigen algorithmischen Autonomie zu kämpfen hat, erzeugt die digitale Hybridität Risiken, die mit den emergenten Eigenschaften kollektiven Verhaltens zusammenhängen. Hier spielen die Sozialwissenschaften ihre Vermittlerrolle zwischen Informatik und Haftungsrechtsdogmatik, indem sie die Bedingungen bestimmen, unter denen sich in sozialen Praktiken Mensch-Maschine-Assoziationen konstituieren.

Weil sie sich bedingungslos dem methodologischen Individualismus verpflichtet fühlen, bleiben ökonomische Autoren allerdings skeptisch gegenüber dem Realitätsstatus von Kollektivakteuren. Sie begreifen diese als bloße „nexus of contracts“ und sehen ihre Personifizierung bestenfalls als eine Abkürzung für komplexe inter-individuelle Beziehungen und schlimmstenfalls als gefährliche „errors“, „traps“ oder „fictions“ (Jensen und Meckling 1976; polemisch zugespitzt Easterbrook und Fischel 1989). In der Welt der Ökonomie zählt nur das Verhalten von Individuen. Im Gegensatz dazu konzentrieren sich soziologische Analysen gerade auf die unterschiedlichen Sozialformationen, die in den Kontakten zwischen Mensch und Algorithmus entstehen (z.B. Nassehi 2019, 224; Hepp 2020; Latour 2007, S. 273 ff.). Diese reichen von kurzfristigen und losen Individual-Interaktionen über Netzwerke bis hin zu kollektiven Mensch-Algorithmus-Assoziationen, die mit interner Arbeitsteilung,

Ressourcenausstattung und Kompetenzverteilung ausgestattet sind.[5] In solchen Konstellationen verdichteter Interaktion, in denen die Einzelhandlungen dem gesamten hybriden Gebilde, nicht aber mehr den einzelnen beteiligten Algorithmen oder Menschen zugerechnet werden (z.B. Pettit 2007), muss das Haftungsrecht auf das Risiko der neuartigen kollektiven Handlungsfähigkeit der Hybride reagieren. Hier sind Normen einer Kollektivhaftung notwendig, die freilich unterhalb der Schwelle der Haftung einer vollwertigen juristischen Person angesiedelt sind.

2.2.3 Digitale Interkonnektivität

Im Gegensatz zu den beiden bisher diskutierten sozio-digitalen Institutionen ist das kollektive Maschinenverhalten ein rein digitales Verhalten, das von der Interkonnektivität autonomer Algorithmen erzeugt wird. Ihr Risiko besteht darin, dass anders als bei digitaler Assistenz kein einzelner handelnder Akteur identifiziert werden kann. Ebenso wenig lässt sich das Risiko der Interkonnektivität mit den Risiken der hybriden Mensch-Maschine-Assoziationen vergleichen. Hier geht es um miteinander vernetzte algorithmische Operationen, denen die gesellschaftliche Kommunikation letztlich ausgesetzt ist, ohne sie zu verstehen. Beim kollektiven Maschinenverhalten findet keine wechselseitige Kommunikation zwischen Menschen und Algorithmen statt, nur eine indirekte strukturelle Kopplung. Die algorithmischen Kalkulationen sind interdependent und lassen sich als ein „restless collective“ qualifizieren, das mit Formen distributiver Kognition arbeitet (Wiedemann 2014, S. 313). Die kollektive Handlung wird dabei durch das interdependente Handeln einzelner Akteure hervorgerufen (Rauer 2017, S. 440 ff. mit Bezug auf Hutchins 1995). Eine solche ‚Kollektivität ohne Kollektiv‘ kann nicht als ein planmäßig organisiertes Sozialgebilde beschrieben werden, weder als Organisation noch als Netzwerk. Es ist ein zufällig entstandener „Schwarm“ von Algorithmen. Das Verhältnis der Gesellschaft zu solchen algorithmischen Schwärmen beschreibt die Systemtheorie als Sozialkontakt zu „unsichtbaren Maschinen“ (Luhmann 1997, S. 117 ff.; ähnlich Hildebrandt 2015, S. 40). Ihr Einfluss auf die Gesellschaft ist nur schwer zu fassen. Weder gibt es hier genuine Kommunikation zwischen Menschen und einzelnen Algorithmen, noch bildet sich ein kommunikationsfähiges Kollektiv aus Menschen und Algorithmen heraus. Statt eines direkten über Kommunikation vermittelten Einflusses üben vernetzte Algorithmen einen nur indirekten über strukturelle Kopplung vermittelten Einfluss auf die sozialen Beziehungen aus. Deswegen dürfte auch eine rechtliche Haftung für die Entscheidungen einzelner Algorithmen ausgeschlossen sein. Wir plädieren stattdessen für Fondslösungen, die politische Entscheidungen von Regulierungsbehörden erfordern, um den je beteiligten Industriesektoren die Verantwortung aufzuerlegen.

2.3 Haftungssubjekte

Die Unterschiede zwischen sozio-digitalen Institutionen und ihren Risiken wirken sich auch auf die Frage aus, welche Akteure oder Organisationen für algorithmische Fehlentscheidungen verantwortlich sind. Auch hier schlagen wir eine differenzierte Behandlung je nach institutionellem Kontext vor. Um die Unterschiede

[5] Die ökonomische Transaktionskostentheorie von Williamson (1996) arbeitet zwar ähnlich mit einer gleitenden Skala von „spot-contracts“ bis hin zu „formal organizations“, verfehlt aber gerade deren korporativen und kollektiven Elemente.

vorab kurz zu kennzeichnen: Auf Entscheidungen einzelner Algorithmen findet die Prinzipal-Gehilfen-Haftung Anwendung, welche die Verantwortung für algorithmisches Fehlverhalten ausschließlich dem menschlichen Nutzer (oder der Organisation) als dem Prinzipal auferlegt, weil er das Risiko der Delegation gesetzt hat. Dagegen haftet für die kollektiven Handlungen der Mensch-Maschine-Hybride das ganze Netzwerk der beteiligten Akteure, also Hersteller, Programmierer, Händler und Nutzer. Interkonnektivitätsschäden schließlich müssen von kollektiven Fonds ersetzt werden, die von der gesamten betroffenen Branche einzurichten sind.

2.3.1 Nutzer/Betreiber

Die digitale Assistenz erzeugt Verantwortlichkeiten nur innerhalb der bilateralen Beziehung zwischen Algorithmus und dem menschlichen Nutzer/Betreiber (oder der Organisation) . Eine Prinzipal-Agent-Beziehung erfordert, dass der Agent die Ziele des Prinzipals verwirklicht und der Prinzipal für die von ihm veranlassten Handlungen des Agenten einsteht. Die Prinzipal-Gehilfen-Haftung zielt gerade nicht auf die Vielzahl der am Computereinsatz beteiligten Akteure – Programmierer, Hersteller, Händler, Betreiber –, sondern ausschließlich auf den Nutzer, der eine Aufgabe an die digitale Technik delegiert und damit das Risiko autonomer Entscheidungen des Algorithmus übernimmt. Für dieses Delegationsrisiko ist daher nur der menschliche Nutzer/Betreiber (oder die Organisation) verantwortlich.

Demgegenüber argumentieren einige Autoren, dass dies die Risiken in ungerechter Weise allein auf den Nutzer/Betreiber verlagert. Sie sehen auch andere Akteure in der Rolle des verantwortlichen Prinzipals, insbesondere den Hersteller, einschließlich des Backend-Betreibers, der im Hintergrund Programm-Updates und ähnliche Leistungen erbringt (Sanz Bayón 2019, unter 7; Europäisches Parlament 2020, Abs. 8). Damit ignorieren sie jedoch die spezifischen Risiken der Aufgabendelegation und gelangen ihrerseits zu einer ungerechten Risikoverteilung zwischen Hersteller und Nutzer. Denn diese setzen ganz unterschiedliche Risiken. Die Risiken müssen jeweils genau definiert und dann ausschließlich denjenigen Akteuren zugewiesen werden, die sie tatsächlich eingegangen sind. Die Prinzipal-Gehilfen-Haftung reagiert auf die Risiken der Arbeitsteilung zwischen Nutzer und Algorithmus. Im Gegensatz dazu reagiert die weiterhin anwendbare Produkthaftung auf die besonderen Risiken der Programmierung, Herstellung und Überwachung der Algorithmen.

2.3.2 Netzwerkmitglieder

Im Gegensatz zu dieser ausschließlich den Nutzer/Betreiber treffenden Prinzipal-Gehilfen-Haftung werden im Fall digitaler Hybridität rechtswidrige Handlungen nicht einzelnen Akteuren, sondern der Mensch-Maschine-Assoziation selbst zugerechnet. Dann aber wird die daraus resultierende finanzielle Haftung auf eine Vielzahl von Akteuren kanalisiert, die ‚hinter' dem digitalen Hybrid stehen. Ein ganzes Netz verschiedener Akteure ist an der Mensch-Maschine-Assoziation beteiligt und profitiert davon. Da die Kontrolle im Netz über die Netzwerkknoten verstreut ist, folgt auch die Verantwortung dieser spezifischen Risikostruktur. Für Mensch-Maschine-Assoziationen lässt sich eine ausgereifte Unternehmenshaftung, die die Assoziation als juristisch handelnde und haftbare Person versteht, (zumindest vorerst) nicht etablieren. Hierzu fehlen dem Recht bisher die dogmatischen Strukturen. Daher

halten wir stattdessen die Prinzipien einer ‚Netzwerkhaftung' für gut geeignet, um die Verantwortung der Beteiligten für die Handlungen der digitalen Hybride zu konstituieren (Vladeck 2014, S. 149; Allain 2013, S. 1074). Die von uns vorgeschlagene digitale Netzwerkhaftung ist am haftungskanalisierenden Modell der amerikanischen „enterprise liability" und an Elementen der deutschen Gesamthandshaftung orientiert. Sie funktioniert in zwei Schritten: erst Handlungszurechnung, dann Haftungszurechnung. Im ersten Schritt wird, ohne dass die Beiträge von Menschen und Algorithmen im Einzelnen identifiziert werden müssen, das Fehlverhalten dem Hybriden als kollektive Handlung zugerechnet. Im zweiten Schritt wird die Haftung für die kollektive Handlung auf die Netzmitglieder kanalisiert. Diese Mitglieder haben das hybride Netzwerk aufgebaut und kontrollieren es, wenn auch nur indirekt. Sie profitieren von dessen Aktivitäten. Als Folge dieser Zurechnung haften alle Netzknoten entsprechend ihrem Anteil, wobei der Anteil anhand des jeweiligen ökonomischen Nutzens und der Kontrolle über das Hybrid bestimmt wird. In Anlehnung an die bekannte Marktanteilshaftung (market share liability) schlagen wir eine ‚Netzanteilshaftung' vor (allgemein dazu: Teubner 2004, S. 246ff., S. 248.). Einzig in dem Fall, dass ein Unternehmen das Netzwerk zentral durch vertragliche Verabredungen koordiniert, sollte die primäre Haftung bei diesem liegen und entsprechend berechnet werden (Loo 2020, S. 189). In der Regel wird dies der Produzent sein, der dann bei den anderen Netzwerkknoten Rückgriff nehmen kann.

2.3.3 Industriesektor

Im Falle der Interkonnektivität ist der Kreis der Haftungssubjekte wieder anders zu ziehen. Die Verantwortung verlagert sich auf ein größeres soziales Kollektiv. Die algorithmische Vernetzung der „unsichtbaren Maschinen" macht es von vornherein unmöglich, einen individuell verantwortlichen Akteur zu bestimmen. Angesichts nur indirekter „struktureller Kopplung" zwischen algorithmischer Interkonnektivität und Gesellschaft kann auch keine Eins-zu-eins-Verantwortungsbeziehung hergestellt werden. Deshalb schlagen wir vor, dass für Schäden aus Interkonnektivität Haftungsfonds einzurichten sind, die von dem beteiligten Industriesektor finanziert werden (Erdélyi und Erdélyi 2020). Die Beiträge der jeweiligen Akteure werden aufgrund ihres Marktanteils und ihrer spezifischen Problemlösungskapazität berechnet. Ein bewährtes Vorbild ist der US-amerikanische Superfund für Umweltschäden. Der Superfund ist nicht nur auf die Entschädigung einzelner Betroffener ausgerichtet, sondern zugleich auf die Kompensation der umfassenderen sozialen und ökologischen Schadensfolgen. Dies sollte auch für Vernetzungsschäden aufgegriffen werden. Dazu müssen Restitutionsmaßnahmen als zusätzliche Instrumente des Haftungsrechts entwickelt werden. Bei Schäden großen Ausmaßes sollte die für den Fonds zuständige Behörde ermächtigt werden, die Akteure auszuwählen, die über eine starke Problemlösungskapazität verfügen, und ihnen die Aufgabe der Restitution und Folgenbeseitigung aufzuerlegen. Die beteiligten Akteure sollten zu Maßnahmen verpflichtet werden, welche die negativen externen Effekte der Interkonnektivität für die Zukunft begrenzen oder sogar ausschalten.

2.4 Rechtsstatus der Algorithmen

Schließlich erfordern die neuen sozio-digitalen Institutionen – digitale Assistenz, Hybridität, Interkonnektivität –, den Rechtsstatus der Algorithmen differenziert zu betrachten. Es kann nicht um eine vollständige rechtliche

Personifizierung von Software-Agenten, Mensch-Computer-Assoziationen oder Multi-Agenten-Systemen gehen. Stattdessen schlagen wir vor, als Antwort auf die drei Risiken den rechtlichen Status von Algorithmen darauf auszurichten, welche Rolle sie jeweils innerhalb ihrer sozio-digitalen Institution spielen.

2.4.1. Teilrechtsfähigkeit

Nur im Falle digitaler Assistenz ist es angemessen, den Algorithmen den Status einer begrenzten Rechtspersönlichkeit einzuräumen. Diese Teilrechtsfähigkeit ermöglicht es ihnen, als Stellvertreter mit bindender Wirkung für ihre Prinzipale Verträge mit Dritten zu schließen. Als Erfüllungs- oder Verrichtungsgehilfen sind Algorithmen mit der erforderlichen rechtlichen Handlungsfähigkeit auszustatten, so dass für ihr Fehlverhalten die Prinzipale haften.

2.4.2. Mitgliedschaft im Hybrid

Im Unterschied dazu besteht die angemessene Antwort auf das Assoziationsrisiko der digitalen Hybridität darin, dem beteiligten Software-Agenten den Rechtsstatus eines vollwertigen Mitglieds der Mensch-Maschine-Assoziation zu verleihen. Eine Maximallösung im zukünftigen Recht würde der hybriden Assoziation selbst Handlungen, Rechte und Pflichten sowie die finanzielle Haftung zurechnen – eine Lösung, die im Privatrecht völlig neue Wege der Rechtspersonalität beschreiten würde. Eine Minimallösung im geltenden Recht würde bloß auf den Rechtsbegriff des Vertragszwecks rekurrieren und ihn für die Auslegung von Computererklärungen und die Bestimmung der Rechte und Pflichten der Teilnehmer nutzbar machen. Wir favorisieren demgegenüber eine mittlere Lösung, die eine Analogie zu den Grundsätzen einer Netzwerkhaftung herstellt. Der Algorithmus erhält damit den Status eines Netzwerkknotens.

2.4.3. Teil des Risikopools

Unsere Antwort auf die Risiken der digitalen Interkonnektivität heißt ‚Risikopool'. Das Haftungsrecht muss dessen Grenzen verbindlich festlegen. Das rechtswidrige Verhalten wird dem Pool zugerechnet. Der Rechtsstatus der Algorithmen besteht dann weder in eigener Personalität noch in Mitgliedschaft in einem Hybrid, sondern darin, ein bloßes Teilelement eines digitalen Risikopools zu sein.

3. Drei Haftungsregimes

3.1. Synopsis

Um die Unterschiede zwischen den drei Haftungsregimes, ihre digitalen und sozialen Voraussetzungen und ihre rechtlichen Normierungen zu verdeutlichen, stellen wir hier in tabellarischer Form eine komprimierte Fassung unserer Vorschläge vor. Die (rekursiven) Zusammenhänge von digitalen Maschinenverhalten, sozio-digitalen Institutionen und Haftungsnormen sehen folgendermaßen aus:

Maschinen-verhalten	Sozio-digitale Institution	Emergente Risiken (Neue Akteure)	Haftungsregime	Haftungssubjekte	Rechtsstatus der Algorithmen
individuell	Digitale Assistenz	Digitale Autonomie (Aktanten)	Prinzipal-Agenten-Haftung	Nutzer/Operator • Gefährdung	Beschränkte Rechtsfähigkeit
hybrid	Mensch-Maschine-Assoziation	Kollektivhandeln (Hybride)	Netzwerkhaftung	Netzwerkknoten • Kontrolle • Ökonomischer Nutzen	Mitgliedschaft im Hybrid
kollektiv	Kopplung an digitale Konnektivität	Distributierte Kognition (Schwärme)	Kollektive Fonds	Industriesektor • Marktanteil • Problemlösungs-kapazität	Teil des Risikopools

Tab. 1: Kausale und normative Beziehungen zwischen Maschinenverhalten, sozio-digitalen Institutionen und Haftungsregelungen.

3.2. Haftungsrechtliche Normen

Zur Konkretisierung der Haftungsregimes schlagen wir die folgenden Regeln vor:

3.2.1. Prinzipal-Gehilfen-Haftung

Die Prinzipal-Gehilfen-Haftung für Fehlentscheidungen eines Algorithmus greift ein, wenn (1) ein menschlicher Prinzipal (oder eine Organisation) eine Aufgabe an einen Algorithmus delegiert, (2) die Delegation die Entscheidungsfreiheit des Agenten erfordert, (3) dem Agenten eine fehlerhafte Handlung zuzurechnen ist, die für einen Programmierer weder vorhersehbar noch erklärbar ist, (4) die Handlung gegen eine vertragliche oder deliktische Sorgfaltspflicht verstößt und (5) die Kausalität zwischen Handlung und Schaden nachgewiesen werden kann.

(6) Als Rechtsfolge haftet ausschließlich der Nutzer des Algorithmus als Prinzipal. (7) Der Schadensersatz beschränkt sich nicht auf die engen Entschädigungsgrundsätze der Gefährdungshaftung, sondern folgt den Schadensersatzregeln des Vertrags- und Deliktsrechts, wonach für über Körper- und Sachschaden hinausgehende Schäden Ersatz zu leisten ist.

3.2.2. Netzwerkhaftung

Sofern die Voraussetzungen der Prinzipal-Gehilfen-Haftung nicht vorliegen, greift eine Netzwerkhaftung ein, wenn (1) in der Zusammenarbeit zwischen Mensch und Maschine (2) eine rechtswidrige Entscheidung getroffen wurde und (3) ihre Tätigkeiten so eng miteinander verflochten sind, dass (4) die Fehlentscheidung weder dem Menschen noch dem Algorithmus zugerechnet werden kann und (5) ein Kausalzusammenhang zwischen den individuellen Handlungen und Schaden nicht

hergestellt werden kann, jedoch (6) nachgewiesen werden kann, dass die kollektive Entscheidung den Schaden verursacht hat.

(7) Als Rechtsfolge haften die Teilnehmer des Akteursnetzwerks, d.h. Produzenten, Programmierer, Händler und die menschlichen Teilnehmer innerhalb des Hybriden. (8) Die Netzwerkhaftung trifft primär den Hersteller als Zentrale des vernetzten Unternehmens. (9) Der Hersteller kann die anderen Teilnehmer entsprechend ihrem Anteil am Netzwerk in Regress nehmen. (10) Der Netzwerkanteil wird durch die Kriterien des wirtschaftlichen Nutzens und der Kontrolle innerhalb des Netzwerks bestimmt.

3.2.3. Haftung des Fonds

Sofern weder Prinzipal-Gehilfen-Haftung noch Netzwerkhaftung eingreifen, ist eine Entschädigung (1) nur dann möglich, wenn ein Fonds oder eine Versicherung für die algorithmischen Schäden eingerichtet worden ist. Voraussetzungen für eine Entschädigung sind, dass (2) die miteinander verbundenen algorithmischen Entscheidungen eine vertragliche oder deliktische Pflicht verletzen und (3) zusammen einen (4) Schaden verursachen.

(5) Als Rechtsfolge haftet der Fonds für die Schäden. (6) Über den Schadensersatz entscheidet eine Behörde, die mit der Verwaltung des Fonds beauftragt wird. Die Behörde bestimmt zugleich (7) die den Fonds finanzierenden Akteure des Wirtschaftszweigs ex-ante nach ihrem jeweiligen Marktanteil und (7) die Akteure, die bei einer ex-post Haftung anhand ihrer Problemlösungskapazität herangezogen werden.

4. Diskussion der Ausgangsfälle

Welche Rechtsfolgen ergeben sich nach den vorgeschlagen Haftungsregeln für die drei in der Einleitung vorgestellten Fälle? Wir hatten zunächst festgestellt, dass nach geltendem Recht in allen Fällen eine Haftung ausgeschlossen ist und die Schäden damit allein vom Geschädigten zu tragen sind. Nach unseren Vorschlägen sind die Fälle je einer sozio-digitalen Institution zuzuordnen und den entsprechenden Haftungsregeln zu unterwerfen.

4.1. Robo-Advice: Digitale Aktanten

Rechtlich stellt sich in diesem Fall zunächst die Frage, ob der Anlagevermittler Costa, der die Beratung und Betreuung von Aktientermingeschäften durch Algorithmen anbietet, für eigene Pflichtverletzungen verantwortlich ist. Wenn der Geschädigte Li jedoch nicht nachweisen kann, dass Costa die Pflichten eines Anlagevermittlers verletzt hat, wird der Erfolg seiner Klage davon abhängen, ob die Fehlentscheidungen des autonom handelnden Algorithmus die Haftung begründen können (Wagner und Luyken 2020; Hughes und Williamson 2019; Sanz Bayón 2019).

Der Stopp-Loss-Befehl des Algorithmus war eine rechtswidrige Entscheidung des Algorithmus. Die hier relevante sozio-digitale Institution ist die digitale Assistenz, bei der der Algorithmus als Agent in einer Prinzipal-Agent-Beziehung auftritt. In diesem Zusammenhang hilft die Produkthaftung nicht weiter, da Costa keine

Herstellerpflichten verletzt hat. Eine verschuldensunabhängige Haftung, die erst vom Gesetzgeber angeordnet werden müsste, würde wiederum viel zu weit gehen. Sie würde der finanziellen Haftung Tür und Tor öffnen. Als reine Kausalhaftung würde sie den Auftraggeber für jede Handlung, ob rechtswidrig oder nicht, des algorithmischen Agenten haftbar machen, sofern diese einen finanziellen Schaden verursacht. Somit bleibt nur die Prinzipal-Gehilfen-Haftung als Anspruchsgrundlage übrig (Strzelczyk 2018). Wenn der Prinzipal Costa einen Finanzmaklervertrag mit Li abgeschlossen und seine vertraglichen Pflichten der Portfolioverwaltung an K1 als seinen Erfüllungsgehilfen delegiert hat, haftet er für alle Verstöße gegen die vertraglichen Pflichten, die K1 begangen hat. Nach geltendem Recht besitzt der Algorithmus jedoch die für die Prinzipal-Gehilfen-Haftung zwingend erforderliche Rechtsfähigkeit (noch) nicht. Gerichte können jedoch autonomen Algorithmen Rechtssubjektivität zusprechen, wie sie es in der Vergangenheit bei bestimmten Assoziationen menschlicher Akteure getan haben. Für die Prinzipal-Gehilfen-Haftung reicht es aus, Algorithmen mit einer Teilrechtsfähigkeit auszustatten, nämlich der Fähigkeit, die vertraglichen Pflichten des Auftraggebers zu erfüllen. Unter diesen Voraussetzungen kann die hier vorgeschlagene Prinzipal-Gehilfen-Haftung als Klagegrund für die Haftung von Costa erfolgreich sein.

4.2. Panama-Papers: Digitale Hybride

Wenn der Algorithmus, der die Vielzahl von Dokumenten in Zusammenarbeit mit den Journalisten analysiert hat, seiner Programmierung entsprechend operiert hat und die menschlichen Journalisten ihren Kontrollpflichten nachgekommen sind und nichts von dem Resultat einer falschen Beschuldigung gewusst haben, kann niemand haftbar gemacht werden (Lewis et al. 2019, S. 69; Oster 2018, S. 29 ff.). Hier ist eine Situation der "kollektiven moralischen Verantwortung“ gegeben, in der eine Gruppe eine unrechtmäßige Handlung begeht, obwohl sich die beteiligten Einzelpersonen korrekt verhalten haben (Copp 2007). Der Algorithmus arbeitete wie programmiert und traf Entscheidungen über die Kennzeichnung, Klassifizierung, Auswahl und Aufbereitung von Informationen für die Verwendung durch Menschen wie vorgesehen, und die menschlichen Journalisten überprüften diese Informationen mit der erforderlichen Sorgfalt auf der Grundlage ihres Wissens. Es ist schwierig, eine einzelne unrechtmäßige Handlung zu identifizieren, obwohl die kollektive Arbeit von Algorithmen und Journalisten zu den unrechtmäßigen Anschuldigungen geführt hat. Für diese Fälle ist eine Netzwerkhaftung wie oben geschildert angemessen. Das Fehlverhalten ist der Mensch-Maschine-Assoziation als Kollektiv von Journalisten und Algorithmen zuzurechnen.

Sofern die unrechtmäßige Handlung dem Mensch-Maschine-Hybrid zugeschrieben werden kann, ist es möglich, die finanzielle Haftung auf die Mitglieder des Netzwerks zu kanalisieren. Der Geschädigte kann den zentralen Knotenpunkt des Netzes erfolgreich verklagen. Im Fall des hybriden Journalismus kann dies entweder die kontrollierende Nachrichtenorganisation oder der Hersteller des Algorithmus sein. Im Zusammenhang mit der algorithmischen Nachrichtenverbreitung könnte es das Unternehmen sein, das den Algorithmus herstellt, also ein Nachrichten- oder Social-Media-Unternehmen. Eine solche Haftung würde unabhängig davon gelten, ob es spezielle gesetzliche Vorschriften für die verschuldensunabhängige Haftung von Nachrichtenanbietern oder Plattformen gibt. Innerhalb des Netzwerks würde die

interne anteilige Verteilung der Haftung entsprechend dem wirtschaftlichen Nutzen und der Kontrolle im kollaborativen Netzwerk erfolgen.

4.3. Flash Crash: Digitale Schwärme

Die von den vernetzten Algorithmen im Finanzmarkt verursachten Schäden sind ein Fall des Interkonnektivitätsrisikos. Die schädigenden Operationen der Algorithmen waren hier zwar von außen ausgelöst, etwa durch betrügerisches Verhalten eines einzelnen Händlers, und begünstigt durch die Infrastruktur des Finanzmarktes selbst, auf dem Algorithmen gleichartiger Programmierung zugelassen sind. Die unmittelbaren Ursachen des Zusammenbruchs sind jedoch die interdependenten Operationen vernetzter Algorithmen. Ihre Geschwindigkeit übersteigt bei weitem menschliche Fähigkeiten. Ein menschliches Eingreifen in die schadensverursachenden Operationen war unmöglich. Vorhersehbarkeit und individuelles Verschulden, wie es für Fahrlässigkeit erforderlich wäre, können in einem solchen Kontext nicht verlangt werden.

Für solche Fälle schlagen wir anstelle der vergeblichen Suche nach einem verantwortlichen Akteur, wie sie die U.S.-Finanzverwaltung versucht hat, eine Fondslösung vor. Der Fonds wird unter die Aufsicht der Finanzmarktbehörden gestellt (ähnlich Yadav 2016, S. 1095). Er wird durch eine relativ geringe Marktzugangsgebühr finanziert, die von den Nutzern und Herstellern der Finanzmarktalgorithmen anhand ihres Marktanteils entrichtet wird.

Hier kommt nun das Kriterium der Rechtswidrigkeit ins Spiel. Nicht jeder Flash Crash sollte den Geschädigten den Zugang zum Fonds eröffnen. Ein durch bloße Volatilität der Finanzmärkte ausgelöster Schaden ist nicht per se ein Haftungsgrund. Der Zugang der Geschädigten zum Fondskapital nach einem Flash Crash ist auf diejenigen Fälle zu beschränken, in denen ein Rechtsverstoß durch vernetzte Algorithmen festgestellt werden kann. Anhand dieses Kriteriums lassen sich Schäden, die durch die normale Volatilität der Finanzmärkte verursacht werden, von Crashs unterscheiden, die auf einen Verstoß gegen die Verhaltensregeln am Finanzmarkt zurückzuführen sind.

Die Finanzmarktaufsichtsbehörde sollte zudem befugt sein, die am algorithmischen Hochfrequenzhandel beteiligten Akteure zu Wiederherstellungs- und Präventionsmaßnahmen anzuweisen. In Bezug auf den algorithmischen Handel wäre dies beispielsweise ein „digital clean up“. Die Behörde würde den beteiligten Firmen auferlegen, systemische Änderungen an der Infrastruktur des algorithmischen Handels vorzunehmen. So könnte das Risiko eines algorithmischen Schwarmverhaltens dadurch gemindert werden, dass eine Verlangsamung der algorithmischen Entscheidungen in das System einprogrammiert wird (Yadav 2016, S. 1097 ff.).

5. Ausblick

Diese drei Haftungsregimes liefern einen wesentlichen Beitrag zu einer sich herausbildenden Digitalverfassung (dazu Celeste 2020, S. 24 ff.). Sicherlich geht es auch darum, eine Optimierung von Sorgfaltsstandards und Aktivitätsniveaus, welche die algorithmischen Aktivitäten zu steuern versucht, anzustreben. Doch sind mit

rechtsökonomischen Argumenten die konstitutionellen Ansprüche an Digitalität nicht erschöpft. Denn in der rekursiven Dynamik Technologie/Sozialinstitutionen/ Haftungsrecht vollzieht sich eine wechselseitige Konstituierung von Rechtsnormen, Rechtssubjekten und sozialen Institutionen. Ja, das Auftreten von autonomen Algorithmen als neuartige Quasi-Subjekte stellt das Privatrecht vor eine Verfassungsfrage erster Ordnung, auf die es womöglich mit der rechtlichen Personifizierung von isoliert agierenden Algorithmen und von Mensch-Algorithmus-Assoziationen sowie der Konstitution von Risikopools samt deren haftungsrechtlichen Konsequenzen reagieren muss. Diese Entwicklung vollzieht sich wie gesagt weder nur technikdeterminiert noch ökonomiedeterminiert, sondern hängt zugleich von vielfältigen politischen Entscheidungen über die gesellschaftliche Verantwortung von digitalen Technologien ab. Sofern in dem nur scheinbar technischen Haftungsrecht über die Bekämpfung von Verantwortungsdefiziten, über Chancen und Gefahren der Personifizierung algorithmischer Prozesse, über die Sorgfaltsmaßstäbe für algorithmische Entscheidungen, über den Schutz von durch digitale Fehlentscheidungen gefährdeten individuellen Rechtspositionen, über das Ausmaß von Entschädigungsleistungen und über die notwendige Re-programmierung von risikobehafteten Algorithmen gestritten und entschieden wird, leistet das Haftungsrecht einen nicht unbeträchtlichen Teilbeitrag zu einer sich gerade entwickelnden Digitalverfassung.

BIBLIOGRAPHIE

Allain, Jessica. 2013. From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems. Louisiana Law Review 73: 1049-1079.

Balkin, Jack. 2015. The Path of Robotics Law. California Law Review Circuit 6: 45-60.

Beardsworth, Thomas und Kumar, Nishant. 2019. Going to Court Over Losses When Robot Is to Blame, *Insurance Journal* 7 May 2019.

Beckers, Anna und Teubner, Gunther. 2021. *Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds* Oxford: Hart.

Celeste, Edoardo. 2020. *Digital Constitutionalism: The Role of Internet Bills of Rights*. Dublin: University College Dublin.

Copp, David. 2007. The Collective Moral Autonomy Thesis. Journal of Social Philosophy 38: 369-388.

Diakopoulos, Nicholas. 2019. *Automating the News: How Algorithms are Rewriting the Media*. Cambridge/Mass.: Harvard University Press.

Easterbrook, Frank H. und Fischel, Daniel. 1989. The Corporate Contract. Columbia Law Review 89: 1416-1448.

Erdélyi, Olivia J. und Erdélyi, Gabor. 2020. The AI Liability Puzzle and a Fund-Based Work-Around. AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 50-56.

Esposito, Elena. 2017. Artificial Communication? The Production of Contingency by Algorithms. Zeitschrift für Soziologie 46: 249-265.

Esposito, Roberto. 2022. *Institution und Biopolitik*. Zürich: Diaphanes.

Europäisches Parlament. 2020. Zivilrechtliche Haftungsregelung für künstliche Intelligenz. Entschließung vom 20. Oktober 2020 2020/2012(INL).

Feenberg, Andrew. 2017. *Technosystem: The Social Life of Reason. Cambridge, MA: Harvard University Press, 2017.* Cambridge, MA: Harvard University Press.

Floridi, Luciano und Sanders, J.W. 2011. On the Morality of Artificial Agents, In *Machine Ethics*, Michael Anderson und Susan L. Anderson (ed.), 184-212, Cambridge: Cambridge University Press.

Graber, Christoph B. 2021. How the Law Learns in the Digital Society. Law, Technology and Humans 3: 12-27.

Hauriou, Maurice E. 1933. *Aux sources du droit: Le Pouvoir, l'ordre et la liberté*. Paris: Bloud & Gay.

Hepp, Andreas. 2020. *Deep Mediatization: Key Ideas in Media & Cultural Studies*. London: Routledge.

Hildebrandt, Mireille. 2015. *Smart Technologies and the End(s) of Law*. Cheltenham: Edward Elgar.

Hughes, Benn und Williamson, Russell. 2019. When AI Systems Cause Harm: The Application of Civil and Criminal Liability. Digital Business Law - Blog

Hutchins, Edwin. 1995. *Cognition in the Wild*. Boston: MIT Press.

Jasanoff, Sheila. 2004. The Idiom of Co-Production', In *States of Knowledge: The Co-production of Science and the Social Order*, Sheila Jasanoff (ed.), 1-12, London: Routledge.

Jensen, Michael und Meckling, William H. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. Journal of Financial Economics 3: 306-360.

Karavas, Vagios. 2009. The Force of Code: Law's Transformation under Information-Technological Conditions. German Law Journal 10: 463-481.

Konertz, Roman und Schönhof, Raoul. 2020. *Das technische Phänomen „Künstliche Intelligenz" im allgemeinen Zivilrecht: Eine kritische Betrachtung im Lichte von Autonomie, Determinismus und Vorhersehbarkeit*. Baden-Baden: Nomos.

Krafft, Tobias D., Zweig, Katharina A. und König, Pascal D. 2022. How to Regulate Algorithmic Decision-Making: A Framework of Regulatory Requirements for Different Applications. Regulation & Governance 16: 119-136.

Latour, Bruno. 2004. *Politics of Nature: How to Bring the Sciences into Democracy*. Cambridge/Mass.: Harvard University Press.

Latour, Bruno. 2007. *Eine neue Soziologie für eine neue Gesellschaft: Einführung in die Akteur-Netzwerk-Theorie*. Frankfurt: Suhrkamp.

Lewis, Seth C., Sanders, Amy Kristin und Carmody, Casey. 2019. Libel by Algorithm? Automated Journalism and the Threat of Legal Liability. Journalism & Mass Communication Quarterly 98: 60-81.

Linardatos, Dimitrios. 2019. Künstliche Intelligenz und Verantwortung. Zeitschrift für Wirtschaftsrecht 504-509.

Lohsse, Sebastian, Schulze, Reiner und Staudenmayer, Dirk. 2019. Liability for Artificial Intelligence, In *Liability for Robotics and in the Internet of Things*, Sebastian Lohsse, Reiner Schulze und Dirk Staudenmayer (ed.), 11-26, Baden-Baden/Oxford: Nomos/Hart.

Loo, Rory van. 2020. The Revival of Respondeat Superior and Evolution of Gatekeeper Liability. Georgetown Law Journal 109: 141-189.

Luhmann, Niklas. 1965. *Grundrechte als Institution: Ein Beitrag zur politischen Soziologie*. Berlin: Duncker & Humblot.

Luhmann, Niklas. 1987 [1972]. *Rechtssoziologie*. Opladen: Westdeutscher Verlag.

Luhmann, Niklas. 1997. *Die Gesellschaft der Gesellschaft*. Frankfurt: Suhrkamp.

Muhle, Florian. 2018. Sozialität von und mit Robotern? Drei soziologische Antworten und eine kommunikationstheoretische Alternative. Zeitschrift für Soziologie 47: 147-163.

Nassehi, Armin. 2019. *Muster: Theorie der digitalen Gesellschaft*. Munich: C.H.Beck.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance* Cambridge: Cambridge University Press.

Oster, Jan. 2018. Haftung für Persönlichkeitsverletzungen durch Künstliche Intelligenz. UFITA - Archiv für Medienrecht und Medienwissenschaft 14-52.

Pettit, Philip. 2007. Responsibility Incorporated. Ethics 117: 171-201.

Pratt, Nicholas. 2018. HK Invstor Sues Robo-adviser in Potential Landmark Case about AI Liability. Commercial Risk 14 May.

Rahwan, Iyad, et al. 2019. Machine Behaviour. Nature 568: 477-486.

Rauer, Valentin. 2017. Distribuierte Handlungsträgerschaft. Verantwortungsdiffusion als Problem der Digitalisierung sozialen Handelns, In *Politik und Verantwortung: Analysen zum Wandel politischer Entscheidungs- und Rechtfertigungspraktiken*, Christopher Daase, Julian Junk, Stefan Kroll und Valentin Rauer (ed.), 436-453, Baden-Baden: Nomos.

Reyes, Carla L. 2021. Autonomous Corporate Personhood. Washington Law Review 96: 1453-1510.

Romano, Santi. 1918. *L'ordinamento giuridico*. Florenz: Sansoni.

Sanz Bayón, Pablo. 2019. A Legal Framework for Robo-Advisors, In *Datenschutz / LegalTech*, Erich Schweighofer, Franz Kummer, Ahti Saarenpää und Burkhard Schafer (ed.), 311-318, Bern: Weblaw.

Schmitt, Carl. 1922. *Politische Theologie*. München: Duncker & Humblot.

Strzelczyk, Bret E. 2018. Rise of the Machines: The Legal Implications for Investor Protection with the Rise of Robo-Advisors. DePaul Business & Commercial Law Journal 16: 54-85.

Teubner, Gunther. 2001. Rechtsirritationen: Zur Koevolution von Rechtsnormen und Produktionsregimes, In *Moral und Recht im Diskurs der Moderne: Zur Legitimation gesellschaftlicher Ordnung*, Günther Dux und Frank Welz (ed.), 351-381, Opladen: Leske und Budrich.

Teubner, Gunther. 2004. *Netzwerk als Vertragsverbund: Virtuelle Unternehmen, Franchising, Just in Time in sozialwissenschaftlicher und juristischer Sicht*. Baden-Baden: Nomos.

Trüstedt, Katrin. 2020. Representing Agency. Law & Literature 32: 195-206.

Trüstedt, Katrin. 2022. *Stellvertretung: Zur Szene der Person*. Konstanz: Konstanz University Press.

Unger, Roberto M. 1996. Legal Analysis as Institutional Imagination. Modern Law Review 59: 1-23.

United States Commodity Futures Trading Commission und U.S. Securities & Exchange Commission. 2016. Findings Regarding the Market Events of May 6, 2010. http://www.sec.gov/news/studies/2010/marketevents-report.pdf.

Vesting, Thomas. 2021. *Gentleman, Manager, Homo Digitalis: Der Wandel der Rechtssubjektivität in der Moderne*. Weilerswist: Velbrück.

Vladeck, David C. 2014. Machines without Principals: Liability Rules and Artificial Intelligence. Washington Law Review 89: 117-150.

Wagner, Gerhard. 1996. Transversale Vernunft und der soziologische Blick: Zur Erinnerung an Montesqieu. Zeitschrift für Soziologie 25: 315-329.

Wagner, Gerhard und Luyken, Lina. 2020. Haftung für Robo Advice, In *Festschrift für Christine Windbichler*, Gregor Bachmann, Stefan Grundmann, Anja Mengel und Kaspar Krolop (ed.), 155-176, Berlin: de Gruyter.

Welsch, Wolfgang. 1996. *Vernunft: Die zeitgenössische Vernunftkritik und das Konzept der transversalen Vernunft*. Frankfurt: Suhrkamp.

Wiedemann, Carolin. 2014. Between Swarm, Network, and Multitude: Anonymous and the Infrastructures of the Common. Distinktion: Scandinavian Journal of Social Theory 15: 309-326.

Wielsch, Dan. 2018. Contract Interpretation Regimes. Modern Law Review 81: 958-988.

Williamson, Oliver E. 1996. *Transaktionskostenökonomik*. Hamburg: Lit Verlag.

Winner, Langdon. 1980. Do Artifacts Have Politics? Daedalus 109: 121-136.

Yadav, Yesha. 2016. The Failure of Liability in Modern Markets. Virginia Law Review 102: 1031-1100.